



(12)发明专利

(10)授权公告号 CN 109196487 B

(45)授权公告日 2020.07.14

(21)申请号 201680086074.9

(22)申请日 2016.09.15

(65)同一申请的已公布的文献号

申请公布号 CN 109196487 A

(43)申请公布日 2019.01.11

(30)优先权数据

15/180,806 2016.06.13 US

(85)PCT国际申请进入国家阶段日

2018.11.23

(86)PCT国际申请的申请数据

PCT/US2016/051850 2016.09.15

(87)PCT国际申请的公布数据

WO2017/218025 EN 2017.12.21

(73)专利权人 超威半导体公司

地址 美国加利福尼亚州

(72)发明人 威廉·埃文·琼斯三世

(74)专利代理机构 上海胜康律师事务所 31263

代理人 樊英如 邱晓敏

(51)Int.Cl.

G06F 12/0862(2016.01)

审查员 李婉怡

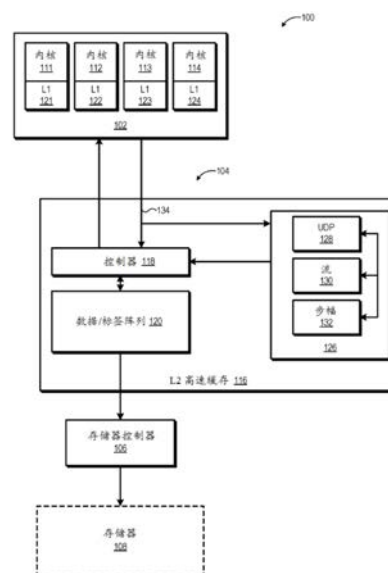
权利要求书3页 说明书10页 附图4页

(54)发明名称

一种在处理系统中预取数据的方法和系统

(57)摘要

在一种包括高速缓存(104)的处理系统(100)中,一种方法包括:监测线程的需求高速缓存存取,以保持所述线程的需求高速缓存存取指向在第一方向上与作为所述线程的一组采样的高速缓存存取的目标的高速缓存线相邻的高速缓存线的次数的第一增减计数。响应于确定第一增减计数已经超过第一阈值(228),所述方法进一步包括:启用第一预取模式(128),在其中,接收到的所述线程的需求高速缓存存取触发对在所述第一方向上与所述接收的需求高速缓存存取锁定的高速缓存线相邻的高速缓存线的预取请求(234)。



1. 一种在包括高速缓存的处理系统中的方法,所述方法包括:

根据采样方案对线程的高速缓存存取进行采样,以确定一组采样的高速缓存存取,其中所述采样方案包括基于所述线程的高速缓存存取的百分比的采样,其中所述采样方案通过从高速缓存未命中比从高速缓存命中更频繁地进行采样,相对于在所述高速缓存处导致高速缓存命中的高速缓存存取,优先处理在所述高速缓存处导致高速缓存未命中的高速缓存存取的采样;

监测所述线程的需求高速缓存存取,以保持所述线程的需求高速缓存存取指向在第一方向上与作为所述线程的所述一组采样的高速缓存存取的所述高速缓存存取的目标的高速缓存线相邻的高速缓存线的次数的第一增减计数;以及

响应于确定所述第一增减计数已经超过第一阈值,启用第一预取模式,其中,所述线程的接收的需求高速缓存存取触发对在所述第一方向上与所述接收的需求高速缓存存取锁定的高速缓存线相邻的高速缓存线的预取请求。

2. 根据权利要求1所述的方法,其中所述第一方向包括下一高速缓存线方向。

3. 根据权利要求1所述的方法,其中所述第一方向包括先前高速缓存线方向。

4. 根据权利要求1所述的方法,其进一步包括:

响应于所述第一预取模式被启用并且响应于接收所述线程的需求高速缓存存取,识别在所述第一方向上与所述需求高速缓存存取锁定的高速缓存线相邻的高速缓存线,并且针对识别的所述高速缓存线发布预取请求。

5. 根据权利要求1所述的方法,其进一步包括:

响应于所述第一预取模式被启用,响应于接收所述线程的需求高速缓存存取,并且响应于所述处理系统的另一预取器未响应于所述需求高速缓存存取而发布预取请求的表示,识别在所述第一方向上与所述需求高速缓存存取锁定的高速缓存线相邻的高速缓存线,并且针对识别的所述高速缓存线发布预取请求;以及

响应于所述第一预取模式被启用,响应于接收所述线程的需求高速缓存存取,并且响应于所述处理系统的另一预取器正响应于所述需求高速缓存存取而发布预取请求的表示,抑制对在所述第一方向上与所述需求高速缓存存取锁定的所述高速缓存线相邻的所述高速缓存线发出预取请求。

6. 根据权利要求1所述的方法,其进一步包括:

监测所述线程的需求高速缓存存取,以保持所述线程的需求高速缓存存取指向在第二方向上与作为所述线程的所述一组采样的高速缓存存取的目标的高速缓存线相邻的高速缓存线的次数的第二增减计数;以及

响应于确定所述第二增减计数已经超过第二阈值,启用第二预取模式,其中,所述线程的接收的需求高速缓存存取触发对在所述第二方向上与所述接收的需求高速缓存存取锁定的高速缓存线相邻的高速缓存线的预取请求。

7. 一种在包括高速缓存的处理系统中的方法,所述方法包括:

基于线程的高速缓存存取的百分比对所述线程的需求高速缓存存取流进行采样:通过从高速缓存未命中比从高速缓存命中更频繁地进行采样,相对于在所述高速缓存处导致高速缓存命中的需求高速缓存存取,优先处理在所述高速缓存处导致高速缓存未命中的需求高速缓存存取的采样;

对于每个采样的需求高速缓存存取,将表示由所述采样的需求高速缓存存取锁定的高速缓存线的地址存储到地址缓冲器的相应条目中;

对于所述线程的多个需求高速缓存存取的每个需求高速缓存存取:

响应于确定所述需求高速缓存存取的地址在第一方向上与存储在所述地址缓冲器的条目中的地址相邻,增加与所述条目相关联的第一计数;以及

响应于确定所述需求高速缓存访问的所述地址在第二方向上与存储在所述地址缓冲器的条目中的地址相邻,增加与所述条目相关联的第二计数;

响应于触发事件,通过对与所述地址缓冲器的所述条目相关联的所述第一计数求和来确定第一总计数并且通过对与所述地址缓冲器的所述条目相关联的所述第二计数求和来确定第二总计数;

响应于确定所述第一总计数已经超过第一阈值,启用第一预取模式,其中,所述线程的接收的需求高速缓存存取触发对在所述第一方向上与所述接收的需求高速缓存存取锁定的高速缓存线相邻的高速缓存线的预取请求;以及

响应于确定所述第二总计数已经超过第二阈值,启用第二预取模式,其中,所述线程的接收的需求高速缓存存取触发对在所述第二方向上与所述接收的需求高速缓存存取锁定的高速缓存线相邻的高速缓存线的预取请求。

8. 根据权利要求7所述的方法,其进一步包括:

响应于所述第一预取模式被启用并且响应于接收所述线程的需求高速缓存存取,识别在所述第一方向上与所述需求高速缓存存取锁定的高速缓存线相邻的高速缓存线,并且针对识别的所述高速缓存线发布预取请求;以及

响应于所述第二预取模式被启用并且响应于接收所述线程的需求高速缓存存取,识别在所述第二方向上与所述需求高速缓存存取锁定的高速缓存线相邻的高速缓存线,并且针对识别的所述高速缓存线发布预取请求。

9. 根据权利要求8所述的方法,其进一步包括:

响应于所述第一预取模式被启用,响应于接收所述线程的需求高速缓存存取,并且响应于所述处理系统的另一预取器未针对所述需求高速缓存存取发布预取请求的表示,识别在所述第一方向上与所述需求高速缓存存取锁定的高速缓存线相邻的高速缓存线,并且针对识别的所述高速缓存线发布预取请求;以及

响应于所述第二预取模式被启用,响应于接收所述线程的需求高速缓存存取,并且响应于所述处理系统的另一预取器未正在针对所述需求高速缓存存取发布预取请求的表示,识别在所述第二方向上与所述需求高速缓存存取锁定的高速缓存线相邻的高速缓存线,并且针对识别的所述高速缓存线发布预取请求。

10. 一种系统,其包括:

高速缓存,其包括用于处理高速缓存存取的高速缓存控制器;以及

向上/向下预取器,其耦接到所述高速缓存控制器,所述向上/向下预取器:

基于线程的高速缓存访问的百分比对所述线程的高速缓存访问进行采样:通过从高速缓存未命中比从高速缓存命中更频繁地进行采样,相对于在所述高速缓存处导致高速缓存命中的高速缓存存取,优先处理在所述高速缓存处导致高速缓存未命中的高速缓存存取的采样;

保持所述线程的需求高速缓存存取指向在第一方向上与作为所述线程的一组采样的高速缓存存取的目标的高速缓存线相邻的高速缓存线的次数的第一增减命中计数;并且

响应于确定所述第一增减命中计数已经超过第一阈值,实施第一预取模式,其中,所述线程的接收的需求高速缓存存取向所述高速缓存控制器触发对在所述第一方向上与所述接收的需求高速缓存存取锁定的高速缓存线相邻的高速缓存线的预取请求。

11. 根据权利要求10所述的系统,其中所述第一方向包括下一高速缓存线方向。

12. 根据权利要求10所述的系统,其中所述第一方向包括先前高速缓存线方向。

13. 根据权利要求10所述的系统,其中所述向上/向下预取器包括:

采样模块,其根据采样方案对所述线程的高速缓存存取进行采样,以确定包括所述采样的高速缓存存取的一组采样的高速缓存存取。

14. 根据权利要求13所述的系统,其中所述向上/向下预取器进一步包括:

地址缓冲器,其耦接到所述采样模块并具有多个条目,所述多个条目被配置成存储与所述采样的高速缓存存取锁定的所述高速缓存线相关联的地址;以及

命中缓冲器,其耦接到所述地址缓冲器并具有多个条目,所述命中缓冲器的每一个条目与所述地址缓冲器的对应条目相关联,并且被配置成存储第一命中计数,所述第一命中计数表示所述线程的需求高速缓存存取已经锁定在所述第一方向上与由存储在所述地址缓冲器的所述相关条目中的所述地址所表示的所述高速缓存线相邻的高速缓存线的次数。

15. 根据权利要求14所述的系统,其中所述向上/向下预取器进一步包括:

命中求和模块,其耦接到所述命中缓冲器,所述命中求和模块基于所述命中缓冲器的所述多个条目的所述第一命中计数的总和来确定所述第一增减命中计数。

16. 根据权利要求15所述的系统,其中所述向上/向下预取器进一步包括:

比较模块,其耦接到所述命中求和模块,所述比较模块响应于所述第一增减命中计数与所述第一阈值的比较来启用对所述第一预取模式的实施。

17. 根据权利要求10所述的系统,其中所述向上/向下预取器进一步将:

监测所述线程的需求高速缓存存取,以保持所述线程的需求高速缓存存取指向在第二方向上与作为所述线程的所述一组采样的高速缓存存取的目标的高速缓存线相邻的高速缓存线的次数的第二增减命中计数,以及

响应于确定所述第二增减命中计数已经超过第二阈值,实施第二预取模式,其中,所述线程的接收的需求高速缓存存取向所述高速缓存控制器触发对在所述第二方向上与所述接收的需求高速缓存存取锁定的高速缓存线相邻的高速缓存线的预取请求。

一种在处理系统中预取数据的方法和系统

技术领域

[0001] 本公开总体上涉及处理系统,且更特别地,涉及在处理系统中预取数据。

背景技术

[0002] 在处理系统中经常使用数据预取来减少与存取存储在存储器中的数据相关联的等待状态或其它延迟。通常,处理系统利用一个或多个预取器来监测数据存取,并根据所监测的数据存取来预测处理系统在不久的将来可能需要哪些数据,并且开始从低级高速缓存或从系统存储器中推测性获取预测数据。常规预取器通常试图识别线程的数据存取模式中的“步幅”或“流”,并利用此步幅或流来预测要请求的下一数据。然而,在许多情况下,这种常规预取器无法准确地检测,通常由于大小限制(例如,预取器可存储的存取“历史”的量)、在给定存储器页面内操作的流数量、或频繁但非总是在本地高速缓存中命中的流而造成的“凝滞”步幅或流模式。因此,常规预取器可能无法识别原本有效的可作为数据推测性预取的适当基础的步幅或流。

发明内容

[0003] 根据第一方面,本发明提供了一种在包括高速缓存的处理系统中的方法,该方法包括:根据采样方案对线程的高速缓存存取进行采样,以确定一组采样的高速缓存存取,其中所述采样方案包括基于所述线程的高速缓存存取的百分比的采样,其中所述采样方案通过从高速缓存未命中比从高速缓存命中更频繁地进行采样,相对于在所述高速缓存处导致高速缓存命中的高速缓存存取,优先处理在所述高速缓存处导致高速缓存未命中的高速缓存存取的采样;监测所述线程的需求高速缓存存取,以保持所述线程的需求高速缓存存取指向在第一方向上与作为所述线程的所述一组采样的高速缓存存取的所述高速缓存存取的目标的高速缓存线相邻的高速缓存线的次数的第一增减计数;以及响应于确定所述第一增减计数已经超过第一阈值,启用第一预取模式,其中,所述线程的接收的需求高速缓存存取触发对在所述第一方向上与所述接收的需求高速缓存存取锁定的高速缓存线相邻的高速缓存线的预取请求。

[0004] 根据第二方面,本发明提供了一种在包括高速缓存的处理系统中的方法,所述方法包括:基于线程的高速缓存存取的百分比对所述线程的需求高速缓存存取流进行采样:通过从高速缓存未命中比从高速缓存命中更频繁地进行采样,相对于在所述高速缓存处导致高速缓存命中的需求高速缓存存取,优先处理在所述高速缓存处导致高速缓存未命中的需求高速缓存存取的采样;对于每个采样的需求高速缓存存取,将表示由所述采样的需求高速缓存存取锁定的高速缓存线的地址存储到地址缓冲器的相应条目中;对于所述线程的多个需求高速缓存存取的每个需求高速缓存存取:响应于确定所述需求高速缓存存取的地址在第一方向上与存储在所述地址缓冲器的条目中的地址相邻,增加与所述条目相关联的第一计数;以及响应于确定所述需求高速缓存访问的所述地址在第二方向上与存储在所述地址缓冲器的条目中的地址相邻,增加与所述条目相关联的第二计数;响应于触发事件,通

过对与所述地址缓冲器的所述条目相关联的所述第一计数求和来确定第一总计数并且通过对与所述地址缓冲器的所述条目相关联的所述第二计数求和来确定第二总计数;响应于确定所述第一总计数已经超过第一阈值,启用第一预取模式,其中,所述线程的接收的需求高速缓存存取触发对在所述第一方向上与所述接收的需求高速缓存存取锁定的高速缓存线相邻的高速缓存线的预取请求;以及响应于确定所述第二总计数已经超过第二阈值,启用第二预取模式,其中,所述线程的接收的需求高速缓存存取触发对在所述第二方向上与所述接收的需求高速缓存存取锁定的高速缓存线相邻的高速缓存线的预取请求。

[0005] 根据第三方面,本发明提供了一种系统,其包括:高速缓存,其包括用于处理高速缓存存取的高速缓存控制器;以及向上/向下预取器,其耦接到所述高速缓存控制器,所述向上/向下预取器:基于线程的高速缓存访问的百分比对所述线程的高速缓存访问进行采样:通过从高速缓存未命中比从高速缓存命中更频繁地进行采样,相对于在所述高速缓存处导致高速缓存命中的高速缓存存取,优先处理在所述高速缓存处导致高速缓存未命中的高速缓存存取的采样;保持所述线程的需求高速缓存存取指向在第一方向上与作为所述线程的一组采样的高速缓存存取的目标的高速缓存线相邻的高速缓存线的次数的第一增减命中计数;并且响应于确定所述第一增减命中计数已经超过第一阈值,实施第一预取模式,其中,所述线程的接收的需求高速缓存存取向所述高速缓存控制器触发对在所述第一方向上与所述接收的需求高速缓存存取锁定的高速缓存线相邻的高速缓存线的预取请求。

附图说明

[0006] 通过参考附图,对本领域技术人员而言,可更好地理解本公开,并且使本公开众多的特征和优点变得显而易见。在不同的附图中使用相同的附图标记来表示相似或相同的项。

[0007] 图1是根据一些实施方案的利用向上/向下预取器的处理系统的框图。

[0008] 图2是示出根据一些实施方案的图1的向上/向下预取器的示例性实施方案的框图。

[0009] 图3是示出根据一些实施方案的图1和图2的向上/向下预取器的操作方法的流程图。

[0010] 图4是示出根据一些实施方案的利用包括图1和图2的向上/向下预取器的一组预取器的预取方法的流程图。

具体实施方式

[0011] 如本文所述,一种处理系统利用向上/向下预取器(UDP)来识别线程的数据存取中的“凝滞”步幅或流模式,并且基于这种被识别的步幅或流模式来产生推测性预取存取。在至少一个实施方案中,UDP试图识别正以相对缓慢上行或下行的方式存取被线程存取的数据组的情况。UDP通过以下来实现这一点:对线程的需求高速缓存存取(针对高速缓存命中和高速缓存未命中两者)进行采样,以及将所采样的需求高速缓存存取锁定的高速缓存线的地址存储在地址缓冲器中。同时,对于UDP观察到的每一个需求高速缓存存取,UDP确定需求高速缓存存取是否寻址到与由地址缓冲器中的地址所表示的高速缓存线相邻的高速缓存线。如果需求高速缓存存取的地址指向在向上方向(即,在较高的存储器地址)上与缓冲

的地址条目的高速缓存线相邻的高速缓存线,则增加与缓冲的地址条目相关联的向上(+)命中计数。同样地,如果需求高速缓存存取地址指向在向下方向(即,在较低的存储器地址)上与缓冲的地址条目的高速缓存线相邻的高速缓存线,则增加与缓冲的地址条目相关联的向下(-)命中计数。

[0012] 在已评估规定数量的需求高速缓存存取之后,或者响应于另一触发事件,UDP将所有向上命中计数器的值相加,并且将所有向下命中计数器的值相加,以分别确定总的向上命中计数和总的向下命中计数。如果总的向上命中计数等于或高于规定阈值,则UDP启用“向上”预取,其中对于线程的每一个需求高速缓存存取,UDP触发对在向上方向上与需求高速缓存存取锁定的高速缓存线相邻的高速缓存线的预取。同样,如果总的向下计数等于或高于规定阈值(该阈值可与向上计数的阈值相同或不同),则UDP启用“向下”预取,其中对于线程的每一个需求高速缓存存取,UDP触发对在向下方向上与需求高速缓存存取锁定的高速缓存线相邻的高速缓存线的预取。在一些实施方案中,UDP与其它预取器协作,使得仅当没有其它预取器响应于需求高速缓存存取而触发预取时,才视情况触发向上预取或向下预取。

[0013] 现参考图1,示出了根据本公开的至少一个实施方案的利用向上/向下预取的处理系统100。处理系统100可由以下各种支持计算的电子装置中的任何一种来实施:诸如服务器、台式计算机、笔记本计算机、平板计算机、支持计算的便携式蜂巢电话(例如,“智能电话”)、支持计算的手表或其它可佩戴物品、个人数字助理(PDA)、游戏机等。在所述示例中,处理系统100包括至少一个计算复合体102、高速缓存层级104和存储器控制器106,存储器控制器106又可耦接到系统存储器108,该系统存储器108与处理系统100的其余部分都是“片上”组件或者被实施为外部组件。

[0014] 计算复合体102包括一个或多个处理器内核,诸如所述示例中的四个处理器内核111、112、113、114。处理器内核可包括中央处理单元(CPU)内核、图形处理单元(GPU)内核、数字信号处理器(DSP)内核或其组合。高速缓存层级104包括一个或多个级别的高速缓存,以对从存储器108存取的数据和临时存储以供一个或多个处理器内核111-114存取的数据进行高速缓存。在该特定示例中,高速缓存层级104包括一组四个级别1(L1)高速缓存121、122、123、124和一个级别2(L2)高速缓存116,其中L1高速缓存121-124中的每一个是用于对应的处理器内核111-114中的一个和计算复合体102的一部分的专用高速缓存,而L2高速缓存116在处理器内核111-114之间共享。应明白,还可以采用更进一步的高速缓存级别,例如级别3(L3)高速缓存。在该示例中,在L2高速缓存级别执行本文所述的向上/向下预取进程。然而,应明白,向上/向下预取进程不限于L2高速缓存级别,而是代以可使用本文提供的准则在诸如L1高速缓存级别或L3高速缓存级别的任何高速缓存级别下实施。

[0015] L2高速缓存116包括控制器118、标签阵列和对应数据阵列(本文中统称为“数据/标签阵列120”)以及预取块126。控制器118进行操作以保持数据/标签阵列120中用于在数据/标签阵列120中在L2高速缓存116下高速缓存的高速缓存线的条目,并且处理来自L1高速缓存121-124和来自存储器控制器106的探测和其它高速缓存存取。预取块126包括一个或多个预取器,以监测L2高速缓存116正在处理的需求高速缓存存取,并适当地触发L2控制器118要处理的推测性预取请求。为此,预取块126包括至少一个向上/向下预取器(UDP)128。在所示示例中,预取块126进一步包括流预取器130和步幅预取器132,其中流预取器

130进行操作以基于常规流模式分析来触发推测性预取,步幅预取器132进行操作以基于常规步幅模式分析来触发推测性预取。

[0016] 作为一般操作性概述,高速缓存层级104进行操作以临时对处理器内核111-114使用的数据进行高速缓存。在一些实施方案中,响应于需求高速缓存存取,即响应于对来自处理器内核111-114中的一个的数据的特定、非推测性请求,对数据进行高速缓存。这种高速缓存需求存取通过高速缓存层级104从低级高速缓存(L2高速缓存116)传播到高级高速缓存(L1高速缓存121-124)。在每一个高速缓存级别下,探测该级别的高速缓存以确定所请求数据的有效副本是否在该级别被高速缓存。如果是,则用信号发送高速缓存命中,并将数据副本提供给探测或请求的源。如果不存在有效副本,则用信号发送高速缓存未命中,并将高速缓存存取传播到下一更高的高速缓存级别,并重复该进程。在高速缓存层级104中不存在所请求的数据的情况下,存储器控制器106被指派从系统存储器108存取所请求的数据,然后从系统存储器108存取的数据被高速缓存在高速缓存层级104中的高速缓存中的一个中。

[0017] 同时,预取块126的预取器128、130、132监测由L2控制器118正在处理的高速缓存存取的流(该流由箭头134表示,因此在本文中也称为“流134”)。预取器128、130、132中的每一个试图检测在计算复合体102处执行的每一个线程的对应存取模式,并且当检测到特定存取模式时,根据检测到的存取模式开始推测性预取请求的发布。在一些实施方案中,预取器128、130、132优先处理或以其它方式协调其预取活动,使得只有预取器128、130、132的子组(例如,只有一个)可触发来自需求高速缓存存取的推测性预取。如上所述,流预取器130和步幅预取器132利用公知的常规预取技术。下文参考图2至图4更详细地描述UDP 128的模式检测和预取开始进程的示例。

[0018] 图2示出了根据至少一个实施方案的处理系统100的UDP 128的示例配置。在所述实施方式中,UDP 128包括采样模块202和一个或多个线程专用向上/向下计数组件203。也就是说,UDP 128可实施向上/向下计数组件203的特定数量的实例,其中每一个实例可用于分配给在计算复合体102处执行的对应线程。因此,如果UDP 128实施向上/向下计数组件203的四个实例,则UDP 128可单独跟踪多达四个线程。采样模块202包括:输入耦接到控制器118的输入,其用于监测L2高速缓存116下的高速缓存存取流134;输入用于接收表示采样方案的采样方案配置数据205的输入,采样模块202通过该采样方案对流134的高速缓存存取进行采样;以及提供存储器地址(或其部分)的输出,所述存储器地址表示高速缓存线针对与采样的高速缓存存取相关联的特定线程锁定到向上/向下计数组件203的实例的每个采样高速缓存存取。

[0019] 向上/向下计数组件203的每一个实例包括向上/向下地址缓冲器204、向上/向下命中缓冲器206、命中控制模块207、命中求和模块208、比较模块210和预取发布模块212。向上/向下地址缓冲器204包括具有多个条目214(例如,十六个条目)的内容可寻址存储器(CAM),其中每一个条目214被配置成存储存储器地址的至少一部分。为了说明,在该示例中,UDP 128基于高速缓存线进行操作,并且假设48位物理地址的位6-47(即,PA[47:6])指代存储器中高速缓存线的数据值的地址(而PA[5:0]指代该高速缓存线内的字节),而(下文描述的)所识别的高速缓存存取的物理地址部分PA[47:6]可被存储在向上/向下地址缓冲器204的对应条目214中。命中求和模块208包括多个条目216,每一个条目216与向上/向下地址缓冲器204的条目214中的对应个条目相关联,并且包括两个子条目:向上命中计数

子条目218,其用于存储存储在对应条目214中的地址的向上命中计数;以及向下命中计数子条目220,其用于存储存储在对应条目 214中的地址的向下命中计数。

[0020] 虽然为便于说明,在图2中示出了命中求和模块208和比较模块 210中的每一个的单个实例,但是每一个向上/向下计数组件203包括命中求和模块208和比较模块210两个实例:一个命中求和模块208 和比较模块210,使用向上命中计数子条目218的向上命中计数来计算和评估“向上”命中的数量;以及一个命中求和模块208和比较模块210,使用向下命中计数子条目220的向下命中计数来计算和评估“向下”命中的数量。在向上/向下命中组件203中实施的两个命中求和模块208中的每一个包括多路复用器222、加法器模块224和结算模块 226。多路复用器222具有每一个耦接到对应子条目218/220的多个输入,以及耦接到加法器模块224的输入的输入。加法器模块224进一步包括耦接到结算模块226的输入的输入和耦接到结算模块226的输出的输入。如下文所述,结算模块226可包括锁存器或其它存储组件,用于存储表示向上/向下命中缓冲器206中表示的向上命中或向下命中的总数量的增多命中计数或减少命中计数。响应于触发事件,比较模块210进行操作以将存储在结算模块226中的增多/减少命中计数与规定阈值228(其可被存储在例如配置寄存器、硬编码等中) 进行比较,并且相应地控制启用/停用信号230。

[0021] 预取发布模块212包括接收启用/停用信号230的输入、接收流 134的高速缓存存取的输入以及基于这些输入而选择性发布预取请求的输出。进一步而言,在一些实施方案中,UDP 128的操作与预取块 126的其它预取器130、132的操作相协调,在这种情况下,预取发布模块212进一步包括用于从这些其它预取器中接收预取状态信息 232的输入,其中预取状态信息232表示这些预取器中的另一个是否已经或将要针对在流134中观察到的需求高速缓存存取而发布预取请求。

[0022] 图3示出了根据一些实施方案的实施图2的UDP 128的操作方法300。如图所示,UDP 128的操作包括至少四个并行运行的独立进程,包括:高速缓存存取监控进程301、命中计数进程302、向上预取控制进程303和向下预取控制进程304。总体而言,进程301、302 表示针对线程重复训练UDP 128,并且进程303、304表示UDP 128 基于在进程301、302期间获得的训练信息关于是否启用或停用线程的向上/向下预取而重复采用的决策进程。

[0023] 高速缓存存取监测进程301表示UDP 128监测高速缓存存取的流134、从中对高速缓存存取进行采样并相应地更新向上/向下地址缓冲器204的进程。因此,在方框311处,采样模块202基于由采样方案配置数据205指定的采样方案从流134中对高速缓存存取进行采样。在至少一个实施方案中,从高速缓存命中和高速缓存未命中两者执行采样,但由于高速缓存未命中倾向于更能反映对推测性预取数据的需要,所以可在采样方案中强调或以其它方式优先处理高速缓存未命中。进一步而言,表示较高的高速缓存级别(例如,L1高速缓存 121-124)的推测性预取的高速缓存存取通常被排除在采样之外。此外,采样的高速缓存存取的百分比可基于诸多考虑,包括向上/向下地址缓冲器204中的条目214的数量、预期由UDP 128检测的存取模式的“比率”等。作为一个示例,采样方案对采样模块202进行配置,以过滤掉表示来自低级高速缓存的推测性预取的所有高速缓存存取,并且过滤掉每八个作为在L2高速缓存116处的高速缓存命中的高速缓存存取中的七个,然后从所得的经过滤高速缓存存取流中,对每 512个高速缓存存取中的一个进行采样。

[0024] 在如此选择了采样的高速缓存存取的情况下,在方框312处,采样模块202将采样

的高速缓存存取锁定的高速缓存线的存储器地址部分PA[47:6]写入向上/向下地址缓冲器204的条目214。在至少一个实施方案中,向上/向下地址缓冲器204以循环(round robin)或先进先出(FIFO)方式进行操作,使得一旦所有条目214都充满时,就选择最旧的条目214来利用新采样的高速缓存存取锁定的高速缓存线的存储器地址部分进行覆写。然而,覆写条目214触发了由方框313 和314所表示的进程,以将与覆写条目214相关联的向上/向下命中计数分别合并到针对线程保持的增多命中计数和减少命中计数中。该进程在下文描述了针对条目生成向上/向下命中计数的进程之后进行描述。在任何事件中,在处理采样的高速缓存存取之后,进程301返回到方框311,以从流134中对下一高速缓存存取进行采样。

[0025] 现转向命中计数进程302,该进程用于对锁定到与向上/向下地址缓冲器204中表示的高速缓存线相邻的高速缓存线的高速缓存存取的数量进行计数。因此,命中控制模块207监测高速缓存存取的流 134。当检测到需求高速缓存存取时,在方框321处,命中控制模块207确定高速缓存存取是否被锁定到与向上/向下地址缓冲器204中表示的高速缓存线中的任何一个相邻的高速缓存线。为执行该进程,命中控制模块207确定需求高速缓存存取锁定的高速缓存线的物理存储器地址部分A,并且在向上/向下地址缓冲器204中执行两次查找:在方框322处对存储器地址部分A-1的第一查找,该存储器地址部分 A-1表示先前与需求高速缓存存取锁定的高速缓存线相邻(即,在向下方向上相邻)的高速缓存线的地址;以及在方框323处对存储器地址部分A+1的第二查找,该存储器地址部分A+1表示紧邻(即,在向上方向上相邻)需求高速缓存存取锁定的高速缓存线的高速缓存线的地址。应注意,虽然为便于说明,将方框322和方框323描述为逐次出现,但应明白,可并行执行由这些框所表示的查找。

[0026] 如方框322所示,如果第一查找命中向上/向下地址缓冲器204 的条目214,则表示存在对紧接着命中条目214中表示的高速缓存线之后的高速缓存线的需求高速缓存存取。也就是说,需求高速缓存存取锁定的是从命中条目214表示的高速缓存线“向上”或紧接在该高速缓存线之后的高速缓存线。因此,作为响应,在方框324处,命中控制模块207使存储在与命中条目214相关联的向上命中计数子条目218中的向上命中计数增加。也就是说,命中控制模块207识别出需求高速缓存存取指向作为从在向上/向下地址缓冲器204中表示的高速缓存线“向上”的一个高速缓存线的高速缓存线,并且作为响应,使向上/向下地址缓冲器204中该高速缓存线的向上命中计数增加1以记录该匹配。或者,如方框323所示,如果第二查找命中向上/向下缓冲器204的条目214,则表示存在对紧接着命中条目214中表示的高速缓存线之前的高速缓存线的需求高速缓存存取。也就是说,需求高速缓存存取锁定的是从命中条目214表示的高速缓存线“向下”或紧接在该高速缓存线之前的高速缓存线。因此,作为响应,在方框 325处,命中控制模块207使存储在与命中条目214相关联的向下命中计数子条目220中的向下命中计数增加。也就是说,命中控制模块207识别出需求高速缓存存取指向作为从在向上/向下地址缓冲器204 中表示的高速缓存线“向下”的一个高速缓存线的高速缓存线,并且作为响应,使向上/向下地址缓冲器204中该高速缓存线的向下命中计数增加1以记录该匹配。在两个查找都未命中的情况下,这表示需求高速缓存存取未指向与向上/向下地址缓冲器204中表示的高速缓存线中的任何一个直接相邻的高速缓存线,因此需求高速缓存存取不影响向上/向下命中计数。

[0027] 如进程302所示,存储在与命中计数子条目218中的向上命中计数表示以下需求

高速缓存存取出现的数量,该需求高速缓存存取锁定从由与向上命中计数子条目218相关联的条目214所表示的高速缓存线“向上”的高速缓存线。存储在向下命中计数子条目220中的向下命中计数相应地表示以下需求高速缓存存取出现的数量,该需求高速缓存存取锁定从由与向上命中计数子条目218相关联的条目214所表示的高速缓存线“向下”的高速缓存线。这样,在存储有效高速缓存线地址的条目214被覆写之前,对于线程,与该高速缓存线相关联的向上命中计数和向下命中计数分别被包括在增多命中计数和减少命中计数中。因此,返回到进程301,当采样模块202将新采样的高速缓存存取的高速缓存线地址存储到向上/向下地址缓冲器204中所识别的条目214时,在方框313处,采样模块202确定有效高速缓存线是否将在识别的条目214中被覆写。如果不被覆写,则没有有效的向上/向下命中计数需要考虑,因此高速缓存线地址可被存储到条目214中而无需进一步的动作。

[0028] 然而,如果在所识别的条目214中已经存储了有效高速缓存线地址,则在方框314处,对有效高速缓存线的覆写可作为命中求和模块208的触发事件来操作,以将与所识别的条目214相关联的子条目218、220的向上命中计数和向下命中计数分别合并到增多命中计数和减少命中计数中。因此,对于向上命中计数,采样模块202用信号将所识别的条目214的索引发送到命中求和模块208的多路复用器222,以用于增多命中计数。作为响应,多路复用器222存取存储在具有相同索引的向上命中计数子条目218中的向上命中计数,并且将存取向上命中计数转发到加法器模块224。然后,加法器模块224将该向上命中计数添加到存储在结算模块226中的增减命中计数的当前值,然后将结果存储到结算模块226中,作为对存储在该结算模块226中的增减命中计数的更新。由命中求和模块208执行相同的处理,该模块相对于减少命中计数和存储在指示索引处的子条目220处的向下命中计数保持减少命中计数。进一步而言,在分别将向上命中计数和向下命中计数添加到增多命中计数和减少命中计数之后,将索引处的子条目218、220重置为零,以将存储在对应条目214中的新的高速缓存线地址。

[0029] 如上所述,进程303和304分别表示基于存储在向上/向下命中求和模块208中的增多/向下命中计数,周期性且选择性启用/停用向上预取和向下预取的进程。进程303表示启用向上预取的进程。因此,在进程301的方框311处已对一些规定数量的高速缓存存取进行采样或者在一些其它的触发事件(例如,倒计时定时器激活)之后,启用“向上”实例的比较模块210。一旦启用,在方框331处,比较模块210从结算模块226接收当前增多命中计数,并且将当前增减命中计数与“向上”高速缓存线命中的规定阈值228进行比较。在至少一个实施方案中,规定阈值228表示线程的向上高速缓存线命中的数量的水印。为了说明,可在启用向上预取之前,将阈值228设置为48次向上命中。可经由建模、经验分析等来选择规定阈值228的特定值。默认情况下,停用向上预取,并且如方框332所示,当增多命中计数小于阈值228时,向上预取保持停用,并且针对另一轮询周期重复进程303。然而,如方框333所示,当增多命中计数上升到或高于阈值228时,已经出现被认为适于启用向上预取的向上高速缓存线命中的最小数量,因此比较模块210配置用于向上预取的启用/停用信号230,以便表示已经为线程启用向上预取。

[0030] 进程304表示启用向下预取的进程,并且以类似于进程303的方式进行。因此,在进程301的方框311处已经对一些规定数量的高速缓存存取进行采样之后,启用“向下”实例的比较模块210。应注意,用于进程304的高速缓存存取的这个规定数量可能不同于用于进程

303的数量。一旦启用,用于“向下”实例的比较模块210开始轮询周期。对于每一个轮询周期,在方框341处,比较模块210从结算模块 226接收当前减少命中计数,并且将当前减少计数与“向下”高速缓存线命中的规定阈值228进行比较。应注意,用于“向下”高速缓存线命中的阈值228可不同于用于“向上”高速缓存线命中的阈值228。如同向上预取,默认情况下,停用向下预取,并且如方框342所示,当减少命中计数小于阈值228时,向下预取保持停用,并且针对另一轮询周期重复进程304。然而,如方框343所示,当减少命中计数上升到或高于阈值228时,已经出现被认为适于启用向下预取的向下高速缓存线命中的最小数量,因此比较模块210配置用于向下预取的启用/ 停用信号230,以便表示已经为线程启用向下预取。

[0031] 图4示出了根据至少一个实施方案的与图3的进程301-304相协调的预取发布模块212的操作方法400。如上所述,UDP 128和预取块126的其它预取器130、132进行操作以检测线程的高速缓存存取中的模式,并且根据这些模式,推测性预测预期线程将需要的数据,并且相应地针对该预测的数据发布推测性预取请求。UDP 128尤其进行操作来检测相对缓慢的移动模式,由此线程移动通过数据块,并顺序地存取高速缓存线,要么按照“向上”顺序(即从较低的存储器地址到较高的存储器地址来存取相邻的高速缓存线),要么按照“向下”顺序(即从较高的存储器地址到较低的存储器地址来存取相邻的高速缓存线)。当检测到这种模式时,如进程303、304所示,UDP 128从检测到的模式的方向适当地启用向上预取或向下预取,并且选择性触发对针对线程所检测到的每一个需求高速缓存存取的相邻高速缓存线的推测性预取。

[0032] 因此,在框402处,随着预取发布模块212处接收到线程的需求高速缓存存取或随着预取发布模块212处对线程的需求高速缓存存取的通知开始方法400。在方框404处,预取发布模块212评估“向上”实例的启用/停用信号230和“向下”实例的启用/停用信号230的状态,以确定是启用了向上预取还是向下预取。如上所述且如方框405 所示,与线程相关联的向上/向下计数组件203保持对命中先前相邻高速缓存线的需求高速缓存存取的数量(减少命中计数)增减计数和对命中下一相邻高速缓存线的需求高速缓存存取的数量(增多命中计数)的增减计数,并且当增多命中计数或减少命中计数中的一个超过其相关联的阈值228时,检测对应的相邻存取模式,并且启用向上预取或向下预取。在至少一个实施方案中,在任何给定时间只可启用两种预取模式中的一种,在这种情况下,抑制了对另一种预取模式的启用。

[0033] 如果既未启用向上预取也未启用向下预取,则在方框406处,预取发布模块212响应于在方框402处接收的需求高速缓存存取不采取任何动作。否则,如果启用了向上预取或向下预取中的一个,则在方框408处,预取发布模块212分析由预取块126的预取器128、130、132交换的预取状态信息232,以确定另一预取器是否将响应于在框 402中接收的需求高速缓存存取而发布预取请求。在至少一个实施方案中,再如方框406所示,UDP 128作为预取器130、132的“备份”预取器进行操作,因此,如果这些预取器中的一个发出信号表示预取器已经或将会发布对需求高速缓存存取的预取请求,则UDP 128通过不采取进一步的动作来抑制对其自身的预取请求的任何发布。

[0034] 或者,如果UDP 128没有相对于预取器130、132而解除优先级,或者如果没有其它预取器正在处理需求高速缓存存取,则在方框410 处,预取发布模块212确定在UDP 128处针对线程启用了哪种类型的预取。如果启用了向上预取,则在方框412处,预取发布模块212

针对紧邻(即,在向上方向上相邻)作为在方框402处接收的需求高速缓存存取的目标的高速缓存线的高速缓存线发布推测性预取请求 234(图2)。相反地,如果启用了向下预取,则在方框414处,预取发布模块212针对先前与作为在方框402处接收的需求高速缓存存取的目标的高速缓存线相邻(即,在向下方向上相邻)的高速缓存线发布推测性预取请求234(图2)。作为在方框412或方框414处发布的预取请求234的结果,相邻高速缓存线被预取到L2高速缓存116,因此如果检测到的相邻存取模式准确,则可用于计算复合体102,因此减少了或消除了当对该相邻高速缓存线的高速缓存存取导致高速缓存未命中并因此将线程置于等待状态直到从系统存储器108中取出相邻高速缓存线时另外可能出现的延迟。这样,作为独立预取器或作为其它预取器的备份的UDP 128可进行操作来识别跨越过多存取(即,过于“凝滞”)以至于无法被常规步幅/流预取器检测到的相邻存取模式,从而允许高速缓存层级104更好地预测计算复合体102的数据存取需求,以便更有效地进行数据存取和处理。

[0035] 在一些实施方案中,上述设备和技术在包括一个或多个集成电路(IC)装置(也被称为集成电路封装或微芯片)的诸如以上参考图 1-4描述的处理系统的系统中实施。电子设计自动化(EDA)和计算机辅助设计(CAD)软件工具可用于设计和制造这些IC装置。这些设计工具通常表示为一个或多个软件程序。一个或多个软件程序包括可由计算机系统执行的代码,以操纵计算机系统来对表示一个或多个 IC装置的电路的代码进行操作,从而执行至少一部分进程来设计或调整制造系统以制造电路。该代码可包括指令、数据或指令和数据的组合。表示设计工具或制造工具的软件指令通常存储在计算系统可存取的计算机可读存储介质中。同样,表示IC装置的设计或制造的一个或多个阶段的代码可存储在相同的计算机可读存储介质或不同的计算机可读存储介质中并可从其中进行存取。

[0036] 非暂时性计算机可读存储介质可包括:在使用期间可由计算机系统存取并向计算机系统提供指令和/或数据的任何非暂时性存储介质或非暂时性存储介质的组合。这种存储介质可包括(但不限于):光学介质(例如,光盘(CD)、数字多功能光盘(DVD)、蓝光光盘)、磁介质(例如,软盘、磁带或磁性硬盘驱动器)、易失性存储器(例如,随机存取存储器(RAM)或高速缓存)、非易失性存储器(例如,只读存储器(ROM)或闪存存储器)、或基于微机电系统(MEMS)的存储介质。计算机可读存储介质可被嵌入在以下系统中:计算系统(例如,系统RAM或ROM)、固定地附接到计算系统(例如,磁性硬盘驱动器)、可移除地附接到计算系统(例如,光盘或基于通用串行总线(USB)的闪存存储器)、或经由有线或无线网络耦接到计算机系统(例如,网络可存取存储(NAS))。

[0037] 在一些实施方案中,上述技术的某些方面可由执行软件的处理系统中的一个或多个处理器来实施。软件包括被存储或以其它方式有形地体现在非暂时性计算机可读存储介质上的一组或多组可执行指令。软件可包括指令和某些数据,当由一个或多个处理器执行时,操纵一个或多个处理器来执行上述技术的一个或多个方面。非暂时性计算机可读存储介质可包括例如磁盘或光盘存储装置、诸如闪存存储器的固态存储装置、高速缓存、随机存取存储器(RAM)或其它非易失性存储器装置等。存储在非暂时性计算机可读存储介质上的可执行指令可以是源代码、汇编语言代码、目标代码或由一个或多个处理器解释或以其它方式可执行的其它指令格式。

[0038] 应注意,并非上述一般性描述中的所有活动或元件都是必需的,特定活动或装置的一部分可能不是必需的,并且除所描述的活动以外,可能会执行一个或多个其它活动,或

者可能包括一个或多个其它元件。更进一步而言,所列出的活动的顺序不一定是其被执行的顺序。而且,已经参考具体实施方案描述了技术方案。然而,本领域的普通技术人员应明白,在不脱离如所附权利要求书所阐述的本公开的范围的情况下,可以进行各种修改和改变。因此,本说明书及附图被认为是说明性的而不是限制性的,并且所有这些修改意在被包括在本公开的范围之内。

[0039] 上文已经关于具体实施方案描述了益处、其它优点和问题的解决方案。然而,可能导致任何益处、优点或解决方案发生或变得更明显的益处、优点、问题的解决方案以及任何特征都不应被解释为任何或所有权利要求的关键、必需或本质特征。此外,上文公开的特定实施方案仅仅是说明性的,因为所公开的主题可以对受益于本文教导的本领域技术人员显而易见的不同但等效的方式进行修改和实践。除了在所附权利要求书中描述的之外,无意在对本文示出的构造或设计的细节进行限制。因此,明显上文公开的特定实施方案可被改变或修改,并且所有这样的变化都被视为处于所公开的主题的范围内。因此,本文寻求如所附权利要求书中所阐述的保护。

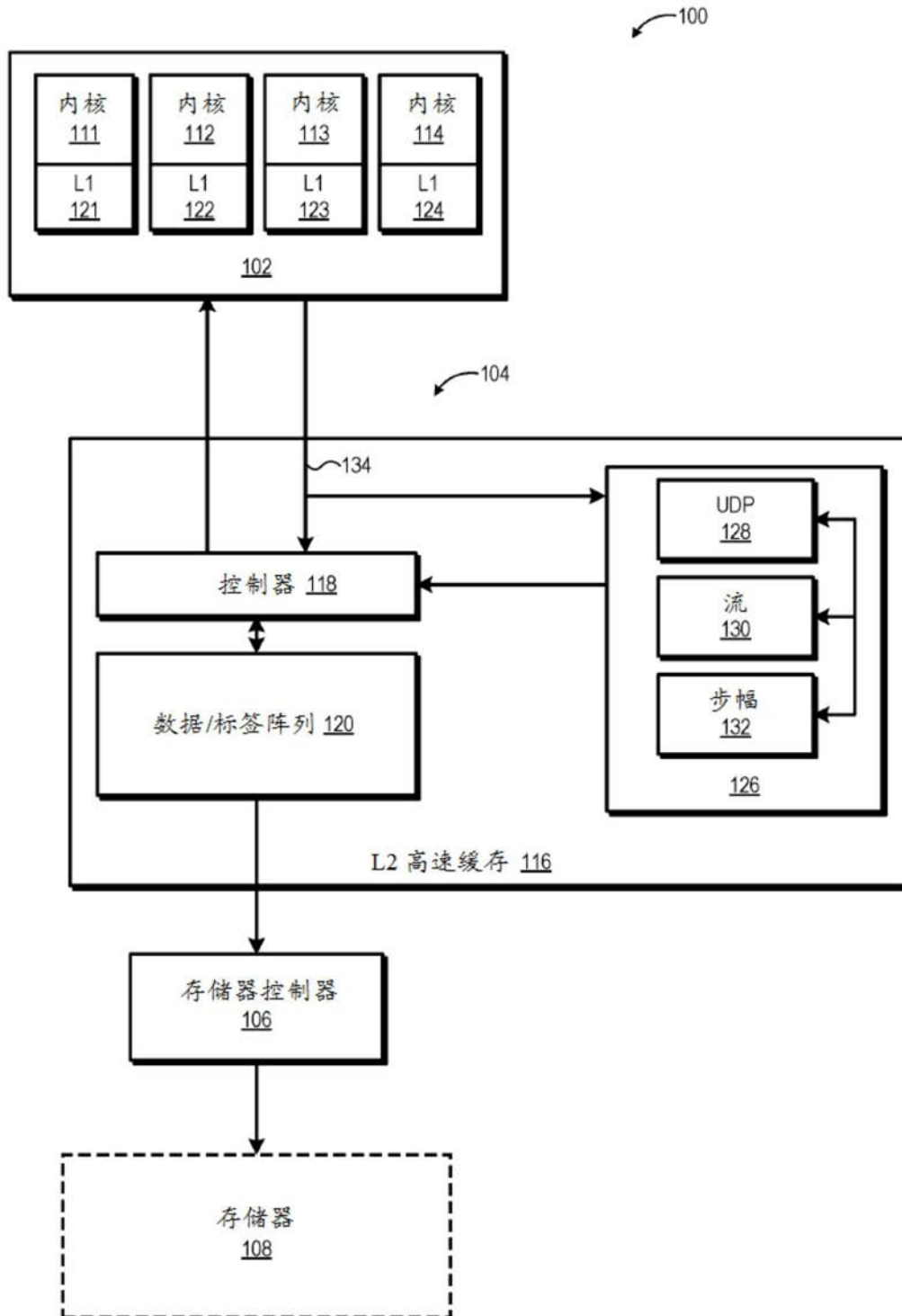


图1

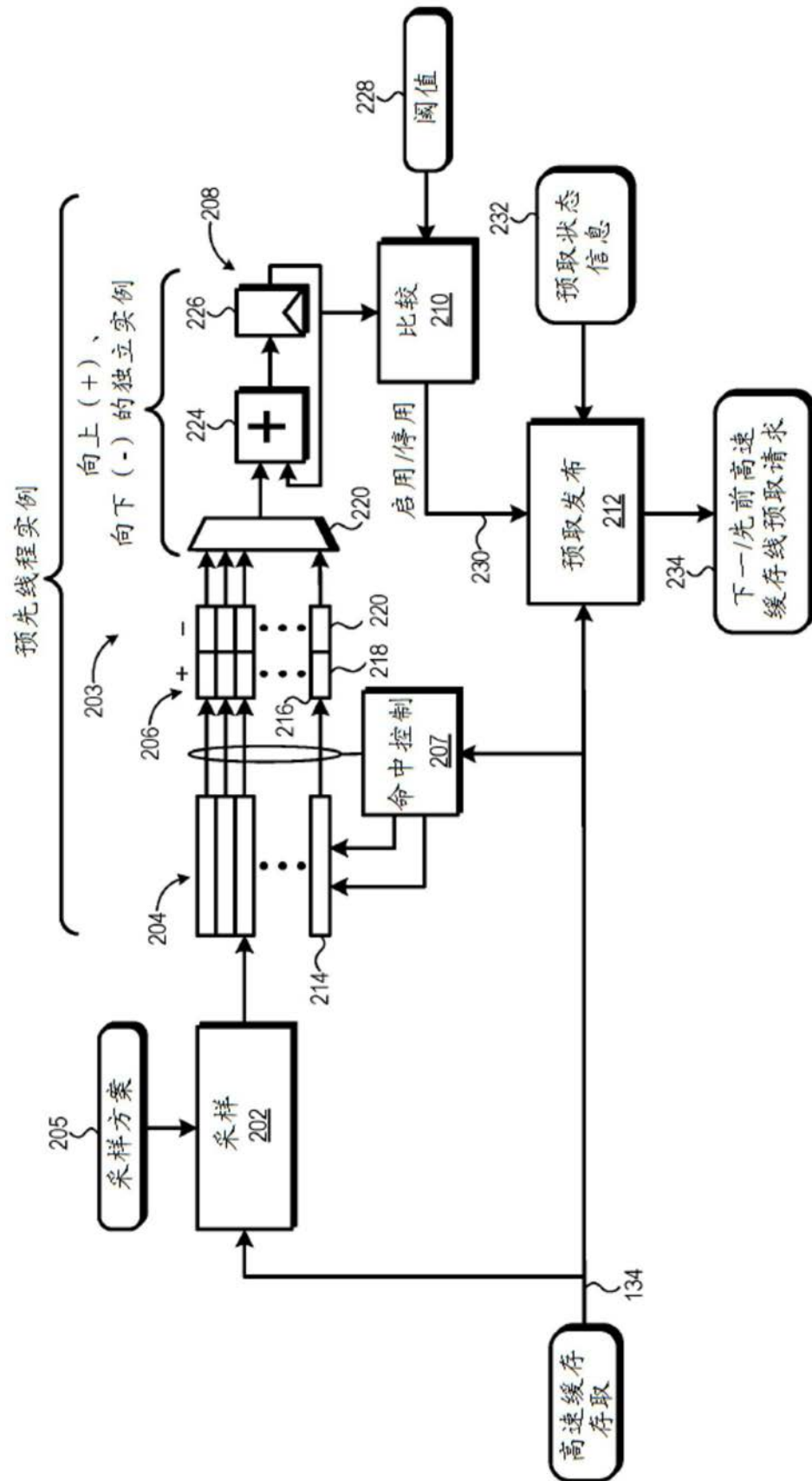


图2

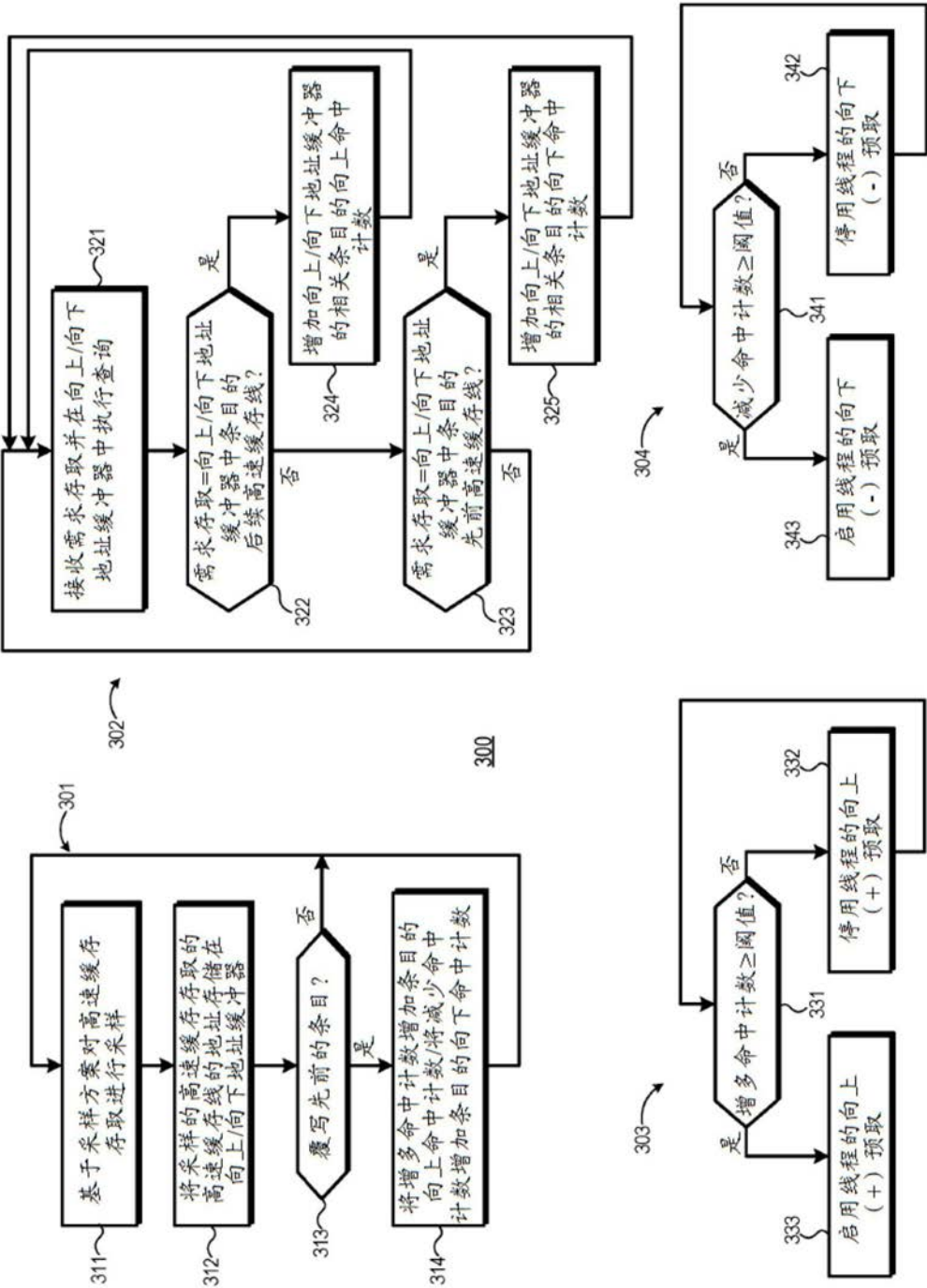


图3

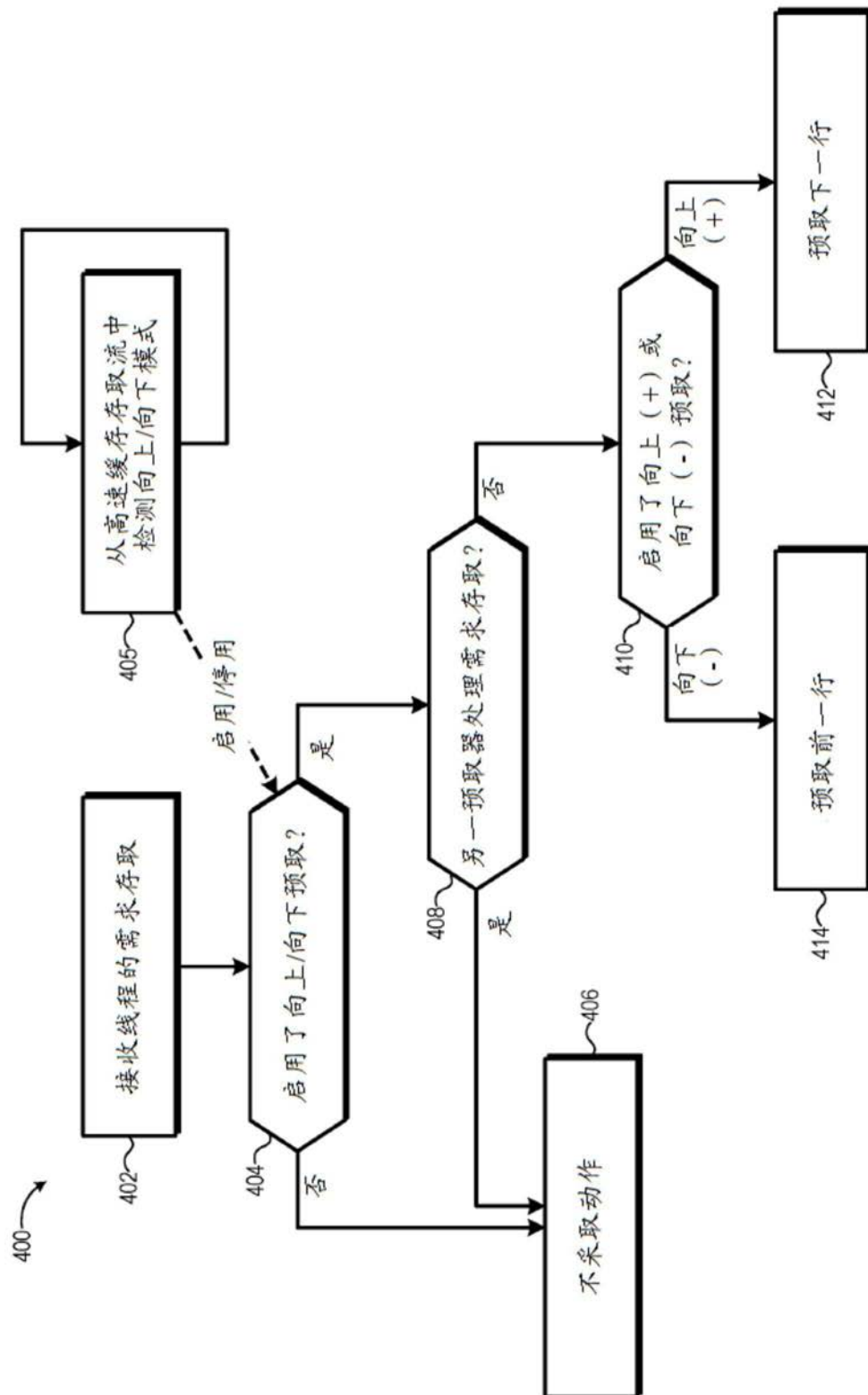


图4